

Servizi di calcolo via web al CILEA: integrazione nel portale del Supercalcolo e aspetti bioinformatici

Mario Marchisio, Paolo Ramieri

CILEA, Segrate

Abstract

Il nuovo servizio di sottomissione job via web sulle macchine del CILEA è ora disponibile sul nostro portale del Supercalcolo. Basato sulla tecnologia Enginframe, tale servizio fornisce, al momento, interfacce grafiche per applicativi di discipline molto diverse fra loro, dalla fluidodinamica alla bioinformatica, e si prefigge di annoverare, in futuro, strumenti per l'esecuzione di programmi su Grid ed aree di lavoro fra gruppi di utenti.

A new service to submit jobs on CILEA's supercomputers is now available at our supercomputing portal. Based on Enginframe technology, it provides the users with some graphic interfaces to run programs belonging to quite different fields, like fluidodynamics and bioinformatics. The possibility to make use of Grid infrastructures and to participate in common working sessions will also be available in the future.

Keywords: Portale, Bioinformatica, Grid.

Un'utile integrazione

Il servizio "Run Your Job" del CILEA è ora disponibile anche tramite il portale per il Supercalcolo www.supercomputing.it.

Si tratta di un'evoluzione molto importante, che mira a far compiere un salto di qualità al nostro portale e a offrire ai nostri utenti una via di accesso più comoda ai servizi di supercalcolo, corredata inoltre da documenti e notizie utili, integrando il tutto in un unico ambiente di lavoro.

"Run Your Job" è un servizio che viene offerto già da qualche anno dal CILEA, anche se fino a ora in veste sperimentale: superata la fase di sviluppo e collaudo del servizio, si entra ora in piena produzione, invitando gli utenti a sfruttare questo nuovo mezzo per accedere alle risorse di calcolo del CILEA.

Come detto, è possibile accedere al servizio dalle pagine del portale del supercalcolo: dalla home page si deve accedere all'area "Supercomputing at CILEA" e da qui al link che porta al servizio "Run Your Job". Si entrerà così in un'area ad accesso riservato da cui, tramite interfacce molto intuitive e di semplice utilizzo, sarà possibile lanciare le proprie analisi sulle

macchine per il supercalcolo del CILEA, in particolare sui cluster Avogadro (con processori Intel Xeon) e Golgi (con processori AMD Opteron) realizzati dalla Exadron [1], divisione HPC della Eurotech S.p.A. [2].

Applicativi disponibili

Il servizio "Run Your Job" è stato realizzato grazie alla tecnologia Enginframe [3] sviluppata dalla Nice Italy [4]. Enginframe ha una struttura di base scritta interamente in Java, può essere usato su qualunque sistema operativo, con qualunque browser e nasconde agli occhi dell'utente finale la complessità degli ambienti di calcolo che andrà a sfruttare. Questo permette al ricercatore che ne fa uso di concentrarsi sul proprio lavoro senza doversi occupare di aspetti più prettamente informatici, quali le modalità di collegamento al cluster, l'utilizzo di sistemi operativi diversi tra loro o il funzionamento dei gestori delle code, ovvero dei software che si occupano di ottimizzare le risorse del sistema di calcolo in funzione delle richieste dell'utenza.

Enginframe, tramite apposite interfacce realizzate con pagine XML e opportunamente

interpretate, permette di inserire i parametri relativi ai calcoli che si intendono svolgere, di indicare quali file di input utilizzare, di monitorare l'andamento della simulazione e di effettuare il download dei risultati direttamente sul proprio PC.



Fig. 1 – Il logo di Enginframe

I servizi attualmente disponibili al CILEA riguardano alcuni applicativi commerciali per la Fluidodinamica Computazionale (FLUENT) e per l'Analisi Strutturale (ABAQUS e RADIOSS) e verranno presto ampliati. Ma il filone di sviluppo più importante riguarda di sicuro il campo della Bioinformatica, per il quale si sta lavorando intensamente. In questo ambito, l'integrazione con il portale per il supercalcolo diventa fondamentale per offrire agli utenti un ambiente di lavoro collaborativo e ricco di strumenti utili.

Gli applicativi bioinformatici

Nel campo della bioinformatica il CILEA partecipa, dall'anno scorso, al progetto LITBIO (Laboratorio Interdisciplinare di Tecnologie Bioinformatiche) insieme con il CNR, Eurotech S.p.A. e le Università di Camerino e di Genova.

Oltre a mettere a disposizione le proprie risorse di calcolo, il CILEA ha il compito di trasferire il vecchio sito web del progetto [5] all'interno del proprio portale, fornendo a una vasta comunità internazionale di bioinformatici (circa 1500) un'interfaccia grafica di facile utilizzo tramite cui accedere a numerosi applicativi comunemente usati nell'analisi di dati di tipo biologico.

Attualmente è possibile effettuare allineamenti di sequenze nucleotidiche o amminoacidiche con BLAST (o, nella sua versione parallela, MPIBLAST) cercando all'interno di database scaricati dal sito dell'NCBI [6] e aggiornati mensilmente.

Più database possono essere interrogati direttamente all'NCBI tramite l'interfaccia di BLASTcl3, mentre una ricerca "in locale" molto veloce è eseguibile tramite BLAT (Blast-like

Alignment Tool). Il quadro sull'allineamento di sequenze è completato dal programma ClustalW, sviluppato per realizzare allineamenti multipli.

L'input di ciascuno di questi programmi può essere sotto forma di file da caricare in locale (cioè dal proprio PC) o in remoto (ovvero dalla propria directory sulle macchine del CILEA) oppure una o più sequenze scritte all'interno di una "textarea". L'output è sempre sotto forma di file di testo, dal momento che ogni job gira tramite un sistema di code.

Di tutt'altra natura è il "workflow" GLUE, un insieme di sei programmi che permette di svolgere analisi di tipo "linkage" sugli alleli di marcatori diversi all'interno di una famiglia di individui, con lo scopo di determinare gruppi di marcatori tra loro correlati. Tipicamente, ciascuno dei sei programmi (Genehunter, Linkmap, Merlin, MLink, Transmit e Unphased) richiede un "pedigree file" e un aggiuntivo file di dati (datafile) da processare. L'output è sia in formato testo che grafico (PostScript o PDF).



Fig. 2 – Il logo del progetto LITBIO

Il futuro del portale

Accanto alle pagine dedicate all'informazione scientifica e alle interfacce web per i vari applicativi che saranno via via installati sui supercomputer del CILEA, il portale dovrà offrire a ogni utente la possibilità di personalizzare il più possibile il proprio account, mettendolo nella possibilità di accedere a interfacce costruite ad hoc, nonché di entrare a far parte di uno o più gruppi di lavoro all'interno dei quali sia possibile lo scambio di file e di informazioni, così come la partecipazione a forum legati a progetti specifici. E' inoltre prevista l'implementazione di un ulteriore servizio che permetta di sottomettere job all'interno delle strutture Grid di cui il

CILEA fa parte. In particolare il CILEA è membro del progetto BioInfoGrid (Bioinformatics Grid applications for life science) [7], che si propone la realizzazione di una infrastruttura Grid europea a supporto delle applicazioni bioinformatiche che richiedono un massiccio uso dei calcolatori: è il caso delle simulazioni di protein folding o di docking molecolare. Quest'ultimo tema, in particolare, risulta essere uno strumento molto utile nella progettazione di farmaci e attrae l'attenzione non solo di svariati centri di ricerca e università, ma anche di numerose aziende farmaceutiche.



Fig. 3 – Il logo del progetto BioinfoGRID

Bibliografia

- [1] URL: <http://www.exadron.com>
- [2] URL: <http://www.eurotech.it>
- [3] URL: <http://www.enginframe.com>
- [4] URL: <http://www.nice-italy.com>
- [5] URL: <http://www.litbio.org>
- [6] URL: <http://www.ncbi.nih.gov>
- [7] URL: <http://www.bioinfoGRID.eu>

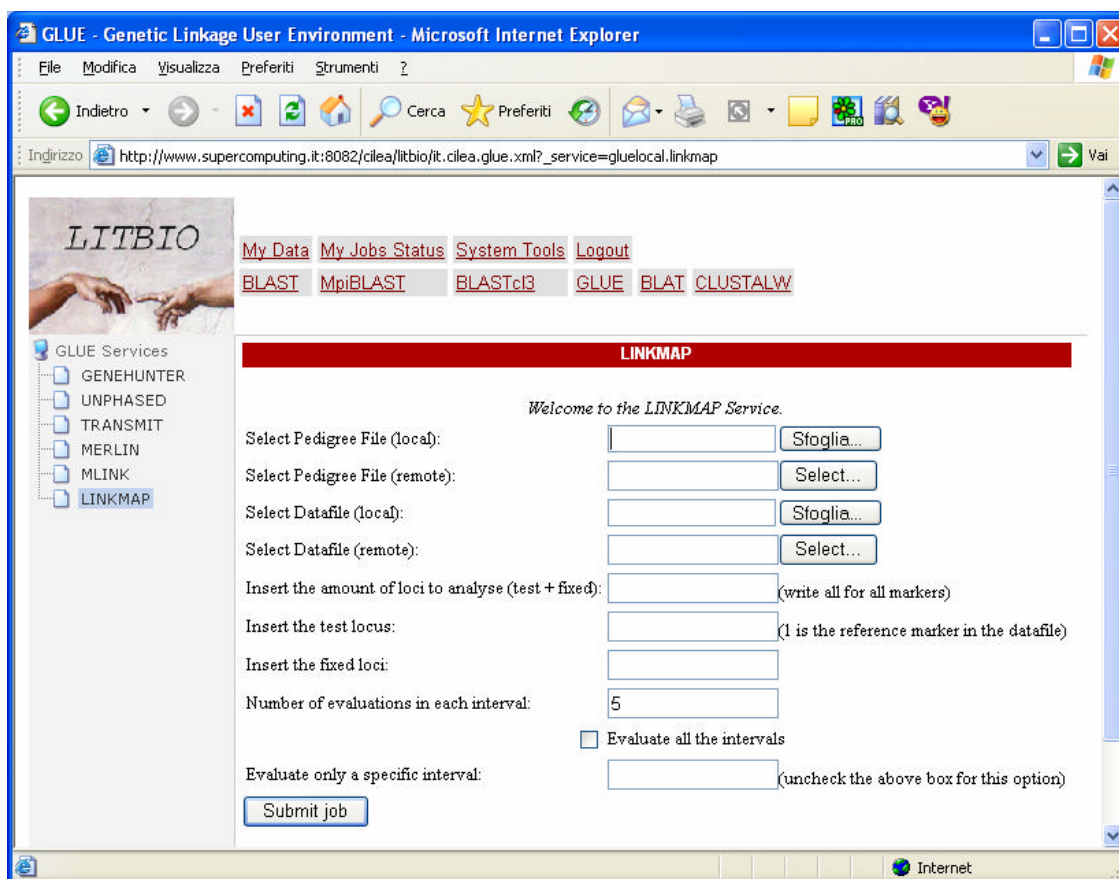


Fig. 4 – Esempio di interfaccia per il software LINKMAP